

ABSTRACT

APPARATUS FOR AND METHOD OF SUMMARISING TEXT

Apparatus for identifying topics of document data has:

5 a word ranker (171) for ranking words that are present in or representative of the content of the document data;

a co-occurrence ranker (172) for ranking co-occurrences of words that are present in or representative of the content of the document data;

10 a phrase ranker (170) for ranking phrases in the document data;

a word selector (174) for selecting the highest ranking words;

15 a co-occurrence identifier (176) for identifying which of the highest ranking co-occurrences contain at least one of the highest ranking words;

a phrase identifier (177) for identifying the phrases containing at least one word from the identified co-occurrences;

20 a phrase selector (178) for selecting the highest ranking one or ones of the identified phrases as the topic or topics of the document data; and

an output device (40) for outputting data relating to the selected topics.

25 (Figs 1 and 6)